Value-at-Risk Estimation for a High-Dimensional Portfolio

A Master Thesis Presented

by

Natalia Sirotko-Sibirskaya

(525754)

to

Prof. Dr. Wolfgang Härdle

Prof. Dr. Ostap Okhrin

CASE - Center of Applied Statistics and Economics

Humboldt-Universität zu Berlin

in partial fulfillment of the requirements

for the degree of

Master of Science in Economics and Management Science

Berlin, November 19, 2013

Contents

List of Figures

List of Tables

Abstract

Increasing complexity of financial instruments and improving data availability along with its variability pose new challenges for the academics as well as for practitioners. In particular, the problem of estimation and inference in case of high-dimensionality receives growing attention. The main difficulty with high-dimensional data lies in the fact that most of the traditional statistical methodology was developed for the case when dimension, p, is lower than the sample size, n. In this thesis several approaches to reduce the dimensionality of the data are analyzed with respect to covariance matrix estimation. In this thesis several approaches to reduce the dimensionality of the data are analyzed with respect to covariance matrix estimation. The estimates of covariance matrix employed in the first part of the thesis are then used for estimation of value-at-risk in the second part of the thesis. For each method a concise theoretical outline is presented. The empirical analysis is performed in Matlab.

Keywords: high-dimensionality, shrinkage, factor models, value-at-risk

1 Introduction

Increasing complexity of financial instruments and improving data availability along with its variability pose new challenges for the academics as well as for practitioners. In particular, the problem of estimation and inference in case of high-dimensionality receives growing attention. The main difficulty with high-dimensional data lies in the fact that most of the traditional statistical methodology was developed for the case when dimension is smaller than the sample size, i. e. when $p < n$, and is not applicable for the case when $p > n$. In order to be able to analyze high-dimensional data statisticians aimed first at dimension reduction. This is based on a general consensus that high-dimensional data admits a lower dimensional representation. To achieve this goal several dimension-reduction methods were developed, for example, feature selection and feature extraction, parsimony and regularization and others.

In this thesis several approaches to reduce the dimensionality of the data are analyzed with respect to covariance matrix estimation. Being a straightforward and easy-to-calculate proxy for analysis of the dependence structure between the variables, sample covariance matrix is applied in various statistical procedures which makes it even more valuable. This is especially true for the area of finance since many financial optimization problems have the covariance matrix of the variables as one of the crucial inputs.

It is a well-known fact that financial data does not satisfy many of the assumption pre-imposed on it in theoretical modeling. One of the established characteristics of financial data is its leptokurtic nature. One observes losses more frequently and they are of a greater magnitude than it is predicted by a standard model. Therefore, the results from the analysis of different covariance matrix estimators are applied for the analysis of various value-at-risk measures. Along with the standard value-at-risk measures a novel risk-measure is introduced, namely, a semi-parametric approach to calculating the VaR which allows to take into account the leptokurtic feature of the data distribution and approximate the losses more precisely.

The goal of the thesis is to analyze performance of different covariance matrix estimators and compare different value-at-risk measures.

The structure of this thesis is the following: first part deals with estimation of a high-dimensional covariance matrix, second part deals with estimation of a value-of-risk of portfolios based on different covariance matrix estimators analyzed in the first part. After a concise theoretical outline of different models and approaches, they are tested empirically on the data. Short description of the data and methodology are available. The empirical analysis is performed in Matlab and the codes are available upon request.

2 High-dimensional covariance matrix estimators

2.1 Literature review

Many empirical problems in finance such as complex portfolio allocation, risk management and asset- and derivative pricing require an estimate of dependence between the variables. Most common way to estimate the dependence between the variables is to use the sample covariance estimator. However, when the sample size is smaller than the dimension, i. e. when $p > n$, such estimator is known to perform poorely as in such a case the sample covariance matrix is not invertible, although the true underlying covariance matrix may exist and be non-singular. Moreover, even if the concentration ratio is less than one but close to it and, thus, the sample covariance matrix is invertible, it can be still numerically ill-conditioned. Therefore, inverting such a numerically ill-conditioned matrix will amplify the estimation error even more and lead to distorted results.

The problem of eigenvalues' distortion can be demonstarted with a simple simulation and analysis of eigenvalues as it is shown on Fig. 1. One can observe that for concentration ratios equal to 1, i.e. $c = p/n = 1$, the curves (and, consequently, corresponding eigenvalues) coincide, whereas for higher dimensions eigenvalues for a sample covariance matrix cannot be calculated - the red curve for $c = 5$ is flat, whereas other curves approximate the true eigenvalues more or less well.

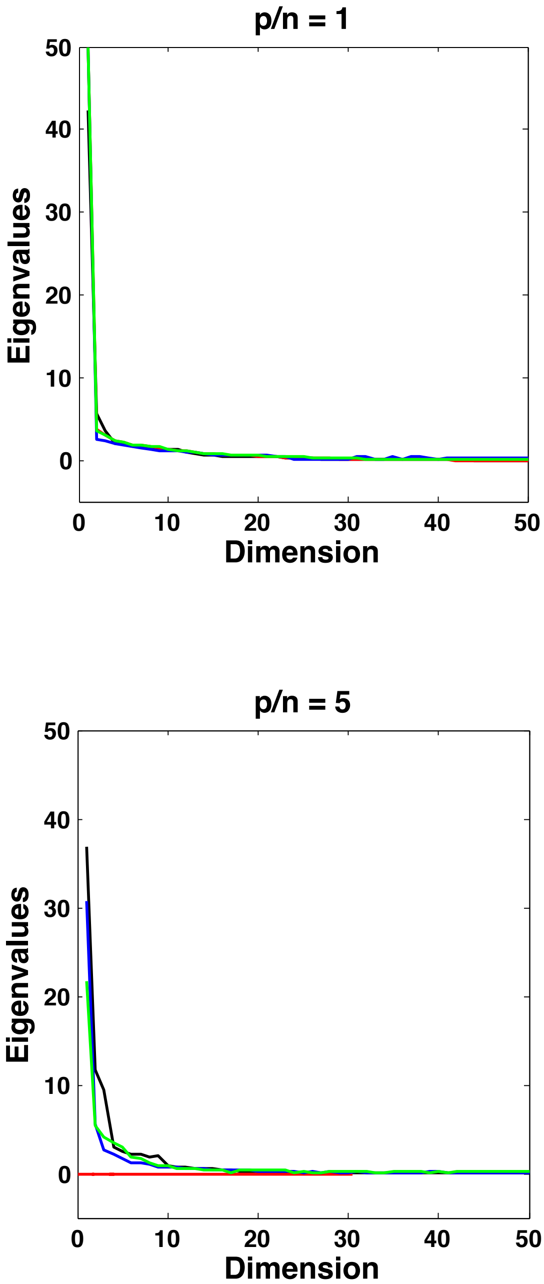


Figure 1: Eigenvalues of a 'true' (black) matrix approximated by estimates from sample covariance matrix estimates (red), factor-based covariance matrix estimates (blue) and a shrinkage-based covariance matrix estimates (green).

Statisticians proposed three approaches to construct a well-conditioned and positive definite estimator of a covariance matrix:

1. find a substitute for a covariance matrix in lower dimensional space using

relevant information;

2. modify the whole sample covariance matrix, for example, use the shrinkage method (indirect modification of sample covariance matrix);

3. perform operations directly on the sample covariane matrix (direct modification of the sample covariance matrix), for example, threshold the eigenvalues forcing them to be positive or use sparsistency assumption.

All of the approaches aim at reducing the initial high dimensionality of the data. However, this aim is achieved in three different ways.

The first approach essentially does not use covariance estimator of initital variables at all. Instead of operating directly with the variables of interest, it uses financial or economic intuition and substitutes the sample covariance estimator of initial variables with the sample covariance estimator of another variables in lower dimensions. Examples of such approach would be Sharpe's single index model **(Sharpe, (1963))** when it is assumed that the covariation between the assets' excess returns is proportional to the variation in the market premium. Another well-established model is that of **Fama and French (1993)** when it is assumed that the variation in excess returns can be explained by the variation in market premium, so called size factor and so called growth factor. Clearly, the key assumption in construction of a factor model requires is identification of economic causality between the variables and existence of measurable data for relevant variables. Therefore, theoretically there can be many models which approximate the volatility of the assets on the market, however, the difficulty lies in choosing either the most robust one so that it can capture the variation in any market or customizing the factors each time so that they are tailored to a specific market. Moreover, the models can be distibguished between dynamic and static, for a more detailed overview see **Bai and Ng (2008)**.

The second approach indirectly modifies the sample covariance matrix, i. e. it shrinks the sample covariance matrix to a certain target. This approach goes back to **James Stein (1956)** who proved that the estimator of individual mean from a normal multivariate distibution can be improved by taking a convex combination of a group mean and a corresponding individual mean. In other words if there are three or more variables of interest coming from a multivariate normal distribution and one is interested in predicting averages for each of them, then pooling-towards-the-mean procedure gives a 'better' result (in statistical sense, 'better' means producing a lower quadratic risk, for example) than simply extrapolating from the three or more separate averages (see **Efron adn Morris (1975)**). The so called James-Stein shrinkage estimator is given as follows: $\bar{x} = \bar{y} + c(y - \bar{y})$, where $\bar{y}$ is the group average, $\bar{x}$ is the individual average and c is the shrinkage intensity. The difficulty with this type of estimator lies in estimating the shrinkage intensity and choosing the correct shrinkage target. The latter difficulty, however, can be seen as well as an advantage since one can choose a specific target, thus, forcing the final estimator to behave in a certain way, e.g. shrinking the covariance matrix towards identity matrix will impose a certain structure on the final estimate, i.e. push the covariance terms to zero. Examples of application of shrinkage method can be found in **Ledoit and Wolff (2003a, 2003b, 2004)**, **Schäfer and Strimmer (2005)**, **Muirhead (1987)**, **Frost and Savarino (1986)**.

The third approach operates directly with the mathematical properties of the sample covariance matrix. Since the goal is to obtain a well-conditioned estimator which means that the eigenvalues of a matrix should be positive, one may threshold the eigenvalues of a matrix (See, for example, **Higham 1988)** and exclude the negative values. Alternatively, one can impose sparsity assumption, or use the penalized likelihood approach (See **Bickel and Levina 2008a,b**).

Often the above-mentioned approaches are combined. For example, the sparsistency assumption is further used in **Fan et al. (2013)** where the sample covariance matrix is approximated by a factor model: the main variation is captured by the principal components and thresholding is applied to the remaining covariance matrix.

There has been a debate about which method is preferrable and provides 'better' results, i. e. 'better' approximates the true covariance matrix than the others. With this respect two of the above mentioned methods, first and second, are often contrasted against each other: factor-based approaches are often being critized that there exists no consensus about which and how many factors should be used. Shrinkage methods are advocated for their robustness. It is interesting, therefore, to compare empirically the performance of factor-based and shrinkage-based methods in order to see whether this critique is justified.

2.2 Theoretical outline

2.2.1 Factor-based estimation

Factor models have been widely used in economics and finance traditionally for forecasting purposes (see **Stock and Watson (2006)**, **Banerjee et al. (2006)**, **Giannone et al. (2007)** and others). In case with high-dimensional covariance matrix estimation the focus is laid rather on using the lower dimensional factor structure to capture the variations in higher dimensions.

Formally the factor model is defined as follows (notations from **Fan et al. (2006)** are adopted):

$$y = B_n f + \varepsilon \tag{1}$$

where $y = (Y_1, ..., Y_p)^T$ are p-dimensional matrix of variables of interest,
$B_n = (b_1, ..., b_p)^T$ with $b_i = (b_{n,i1}, ..., b_{n,iK})^T$, $i = 1, ..., p$, are estimated factor loadings,
$f = (f_1, ..., f_K)^T$, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_p)^T$, are k factors. It is assumed that $(f_1, y_1), ..., (f_n, y_n)$ is n independent and identically distributed sample of (f, y).

Further assumptions are made for the error terms of the model: namely, it is assumed that $E(\varepsilon|f) = 0$ and $\text{cov}(\varepsilon|f) = \Sigma_{n,0}$ is diagonal.

Then the variance of y, Σ_n, is defined as follows:

$$\Sigma_n = \text{cov}(B_n f) + \text{cov}(\varepsilon) = B_n \text{cov}(f) B_n^T + \Sigma_{n,0} \tag{2}$$

The straight-forward way to obtain the estimator for Σ_n is to use the least-squares estimators of B_n, $\text{cov}(f)$ and $\Sigma_{n,0}$. These estimators are obtained as follows:

$$\widehat{B_n} = YX^T(XX^T)^{-1} \tag{3}$$

$$\widehat{\text{cov}}(f) = (n-1)^{-1}XX^T - (n(n-1))^{-1}X11^TX^T \tag{4}$$

$$\widehat{\Sigma_{n,0}} = \text{diag}(n^{-1}\hat{E}\hat{E}^T) \tag{5}$$

$$\tag{6}$$

where $\hat{E} = Y - \hat{B}X$, and X denotes the regressors, or factors.

Therefore, the resulting estimator can be rewritten as follows:

$$\widehat{\Sigma_n} = \widehat{B_n}\widehat{\text{cov}(f)}\widehat{B_n^T} + \widehat{\Sigma_{n,0}} \tag{7}$$

This estimator of the covariance matrix is proven to be asymptotically normal and, most importantly, invertible (See **Fan et al. (2006)**).

There exists different models which can be used in order to construct a better proxy for the variation in the variables of interest. Most prominent ones are Sharpe single-index model (1963), Fama and French 3-factor model (1993), Arbitrage-Pricing theory model by **Ross (1976)** and others. Here two models are tested: single-index model by Sharpe and Fama and French 3-factor model.

Fama and French model is defined as follows:

$$y = b_1 f_1 + b_2 f_2 + b_3 f_3 + \varepsilon_i \tag{8}$$

where $y = (Y_1, ..., Y_p)^T$ are p-dimensional portfolio of stocks, $i = 1, ..., p$, and b_j's are factor loadins, $j = 1, \ldots, 3$. The first factor f_1 denotes the excess return on the proxy of the market portfolio which is a value-weight return of all Center for Research in Security Prices (CRSP) firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11, i.e. ordinary common shares, at the beginning of month t, good shares and price data at the beginning of t, and good return data for t minus the one-month Treasury bill rate (from Ibbotson Associates).

The second factor f_2 is defined as SMB, "small minus big (market capitalization)" is the average returns on the three small portfolios minus the average returns on the three big portfolios:

$$SMB = 1/3(SmallValue + SmallNeutral + SmallGrowth) \tag{9}$$

$$-1/3(BigValue + BigNeutral + BigGrowth) \tag{10}$$

And the third factor f_3 is defined as HML, "high minus low (growth)", is the average return on the two value portfolios minus the average return on the two growth portfolios:

$$HML = 1/2(SmallValue + BigValue) – 1/2(SmallGrowth + BigGrowth) \quad (11)$$

Despite of controversial results obtained when analyzing empirical data, this model is proven to capture the most of the variation in the stock returns. More information on these factors for different portfolios and regions as well as on their construction can be found on Kenneth R. French at his open-source data library: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Sharpe single-index model is defined in the similar manner, however, only the first factor, i.e. market premium is considered to be of relevance. Thus, the resulting equation becomes:

$$y = b_1 f_1 + \varepsilon_i \quad (12)$$

The exact composition of a market portflio is not crucial here, the essense of this factor model lies in the fact that it can be disputed which factors exactly reveal the variation of the stocks better and how many of them should be used, however, it is most likely always to hold that the volatility of the stocks are related to the overall volatility of the market.

Estimators based on the factor models are criticized that although they contain little estimation error, they can be misspecified and biased. Customizing the factors to a specific market turns into the art of correctly choosing those factors which are more relevant for the variables of interest - the process which can require large amount of work, investigation and trial-and-error testing methods. Therefore another method is available that does not require so much pre-knowledge and operates with the sample covariance matrix itself, namely, the shrinkage method.

2.2.2 Shrinkage-based estimation

The shrinkage method goes back to **James Stein's (1956)** idea of estimating an individual average as a convex combination of a pooled average and an individual average. The idea of shrinkage has been widely applied in finance, for example, shrinking method is used in estimating the expected returns, covariance matrices, portfolio weights (see **Golosnoy and Okhrin (2007)**) and others. Also the method of non-linear shrinkage has been developed (see **Ledoit and Wolff (2011, 2013)**).

In this work the focus will be laid on the application of a linear shrinkage methodology developed by **Ledoit and Wolff (2001, 2003, 2004)**. The main advantages of the method by Ledoit and Wolff lies in the absense of computational complexity , in its speed and, essentially, in the absense of distributional assumptions. Moreover, unike the previous literature on the shrinkage method these authors explicitely consider the high-dimensional case, i.e. when $p > n$.

The intuition behind the shrinkage principle is the following: the sample covariance matrix is proven to be statistically unbiased, however, it has large estimation error in high dimensions, whereas the shrinkage target may be severely biased but well-conditioned. The idea is to combine these two estimators in an optimal way to reduce the estimation error and the bias.

$$\Sigma^S = \alpha F + (1-\alpha)S \tag{13}$$

where S is the sample covariance matrix, $\alpha \in (0,1)$ is the shrinkage intensity and F is the shrinkage target. Later the elements of the Σ^S, F and S are denoted as σ_i^s, f_i and σ_i, respectively. The key difficulty lies in choosing an optimal shrinkage intensity α. To make the shrinkage intensity depend on the data the natural idea is to minimize a certain loss function, for example, the mean squared error, to compromise between bias and estimation error. This is done as follows:

$$\begin{aligned}
\mathrm{R}(\alpha) &= \mathrm{E}(\mathrm{L}(\alpha)) && (14)\\
&= \mathrm{E}(\sum_{i=1}^{p}(\sigma_i^s - \sigma_i)^2) && (15)\\
&= \sum_{i=1}^{p} \mathrm{Var}(\sigma_i^s) + (\mathrm{E}(\sigma_i^s) - \sigma_i)^2 && (16)\\
&= \sum_{i=1}^{p} \mathrm{Var}(\alpha f_i + (1-\alpha)s_i) + (\mathrm{E}(\alpha f_i + (1-\alpha)s_i) - \sigma_i))^2 && (17)\\
&= \sum_{i=1}^{p} \alpha^2 \mathrm{Var}(f_i) + (1-\alpha)^2 \mathrm{Var}(s_i) + 2\alpha(1-\alpha)\mathrm{cov}(s_i, f_i) + (\alpha\, \mathrm{E}(f_i - s_i) + \mathrm{Bias}(s_i))^2 && (18)
\end{aligned}$$

Minimization of this loss function, or the risk function of this loss function, i.e. the expectation thereof, provides the estimator of the optimal shrinkage intesity α:

$$\alpha = \frac{\sum_{i=1}^{p} \mathrm{Var}(s_i) - \mathrm{cov}(f_i, s_i) - \mathrm{Bias}(s_i)\,\mathrm{E}(f_i - s_i)}{\sum_{i=1}^{p} \mathrm{E}((f_i - s_i)^2)} \tag{19}$$

Generally, the weight α controls how much structure is imposed: the larger is the weight, the stronger is the structure. However, more observations can be made about the size of the shrinkage intensity:

1. The optimal shrinkage intensity is diminishing in variance of s_i;
2. If the estimation error of S and F are positively correlated, then the weight put on the shrinkage target, F, decreases. In other words, the advantage of using the shrinkage target is reduced;
3. The optimal shrinkage intensity is diminishing in the mean squared distance between F and S.

Moreover, it is important to note that in large samples α may exceed one or even be negative, therefore, to avoid negative shrinkage or overshrinkage, the estimated intensity is truncated as follows: $\alpha = \max(0, \min(1, \alpha))$.

Choice of a shrinkage target is motivated by imposition of a lower dimensional structure on the data. The following shrinkage targets are considered.

1. Shrinkage to an identity matrix
2. Shrinkage to a two parameter matrix
3. Shrinkage to a diagonal covariance matrix
4. Shrinkage to a constant correlation matrix
5. Shrinkage to the market

First two targets, shrinkage to identity and to a two-parameter matrix, are the stringest ones, i.e. they shrink all components of the sample covariance matrix: shrinkage to identity forces the diagonal components to be equal to one and off-diagonal components to be equal to zero, whereas shrinkage to a two parameter matrix forces all elements to have the same variance (average over all variances) and the same covariance (average over all covariances). These shrinkage targets are extremely low-dimensional (the number of parameters to be estimated are 0 and 2, correspondingly), therefore, estimation error is indeed severely reduced, however, the probability that these structures are not appropriate for reflection of actual data is rather high.

A third target, shrinkage to a diagonal covariance matrix, is less stringent and shrinks only off-diagonal elements forcing them to be zero. However, it allows for unequal, or stock-specific variances. So, the number of parameters to estimate are p.

A fourth target, shrinkage to a constant correlation matrix, allows for different variances, but shrinks the off-diagonal elements to a constant correlation coefficient. This target is the most fragile one with respect to the issue of high-dimensionality since it has the largest number of parameters to be estimated. It will be shown later that this estimator happens to perform the worst in the empirical setting.

A fifth estimator combines the additional information and a shrinkage principle. In this way, one can account for market covariance without employing the factor structure. Here the market is not the market premium as it was defined for Fama and French model earlier, but the cross-sectional average across all stocks.

Matlab codes for the shrinkage-based estimators are provided by Ledoit and Wolf on their weibsite: `http://www.ledoit.net/research.htm`.

2.3 Empirical analysis

In this subsection different high-dimensional covariance matrix estimators are tested in empirical setting. The performance of the estimators are compared in terms of their standard deviations, returns and Sharpe ratios.

2.3.1 Portfolio selection

Portfolio selection is based on Markowitz portfolio optimization framework. Consider N stocks with mean μ and a covariance matrix Σ. Then the problem of portfolio selection can be defined as follows:

$$\min_{w} w^T \Sigma w \tag{20}$$

$$\text{subject to} \quad w^T 1 = 1 \tag{21}$$

$$w^T \mu = q \tag{22}$$

where 1 denotes a vector of ones and q is the espected rate of return that is required on the portfolio. The solution to this minimization problem is given as follows:

$$w = \frac{C - qB}{AC - B^2}\Sigma^{-1} + \frac{qA - B}{AC - B^2}\Sigma^{-1}\mu \tag{23}$$

where

$$A = 1^T \Sigma 1, B = 1^T \Sigma \mu, C = \mu^T \Sigma^{-1} \mu \tag{24}$$

The essence of this framework lies in the trade-off between risk and return, i. e. any reduction in risk translates into a higher return and vice versa. Therefore, in order to form an optimal portfolio one is confronted with two problems: estimating the covariance matrix and expected returns. When $p > n$ the sample covariance matrix cannot be inverted or numerically ill-conditioned, one cannot optimize a high-dimensional portfolio. Various shrinkage estimators and estimators based on imposing structure are supposed to amend this problem since the proposed estimators are unlike the sample covariance matrix estimator are invertible.

Empirically one estimates the covariance matrix up to a certain date based on the historical data, then forms a portfolio by obtaining the weights which have to be given to each asset and holds the protfolio until the next rebalancing occurs. Thus one can measure the performance of the covariance matrix estimator after the portfolio has been formed. This is a measure of out-of-sample performance, or of predictive ability.

There is a discussion in the literature whether estimating the expected returns are more important then estimating the covariance matrix, and vice versa. Since the focus of this work is to estimate the covariance matrix more precisely and to analyze its performance, for the empirical estimation of the expected returns the mean over the moving window with size of $n = 100$ is taken.

2.3.2 Data description

The data on stock returns are taken from Datastream data base access to which is provided by the Research Data Center of the Collaborative Research Center 649 at the Faculty of Economics at Humboldt-University of Berlin, Germany. For the empirical studies the stocks traded on the Singapor stock exchange is chosen. All stocks are denominated in the U. S. $. The time period spans from the 26th of June 2003 till the 30th of September 2013. The choice of a period is explained by the availability of the data.

The data on risk factors are provided by Kenneth R. French at his open-source data library: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. One deficiency in using these risk factors lies in the fact that the factors for Asian market are provided unfortunately only on monthly basis, therefore, the factors for the U. S. market are used. Missing values on risk factors' returns are interpolated.

To obtain the first intuition about the data, it is helpful to visualize it and perform basic summary statistics. Fig. 2 shows the kernel density of the returns on the equally weighted portfolio based on 400 stocks. Also Fig. 2. shows the time series of the data. One can notice that an increase in volatility of the equally-weighted portfolio in 2007 is not reflected in the risk factors, however, volatility clustering observed between 2009 and 2010 and a smaller spike around 2011-2012 are repeated in Singapor as well as in American data. Therefore, the data on U. S. factors can be used as (an imperfect) proxy for the Asian market.

Table 1 summarizes the basic statistical information. All the return series have significant kurtosis and a slightly negative skewness, except for SMB factor which posesses a significantly high and unlike the other series positive skewness. The most volatilie factor is market premium, whereas a portfolio, on average, has a quite low volatility. The 'non-normality' of the data is again proven by the Jarque-Bera test which is rejected at 5% significance level. The Dickey-Fuller test (with the null hypothesis being that a process has a unit root) and KPSS test (with the null hypothesis being that a process is stationary) provide contradictory results for the equally-weighted portfolio of 400 STI stocks. lthough, it is known that KPSS test is more stringent, it is more plausible assumption that the returns data are not likely to be stationary for the whole estimation period of 10 years from 2003 to 2013.

Further, the time series are checked for autocorrelation. For the risk factors only very small autocorrelation is observed (less that 0.05 in magnitude), whereas for the time series of each stock autocorrelation is different, however, for almost all the stocks the lag would be not greater than 1. More accurate procedure would suggest fitting, for example, AR(1) process to the time series and working with standardized residuals, however, since only small serial autocorrelation is observed, this step is not considered to be of crucial importance for this particular analysis.

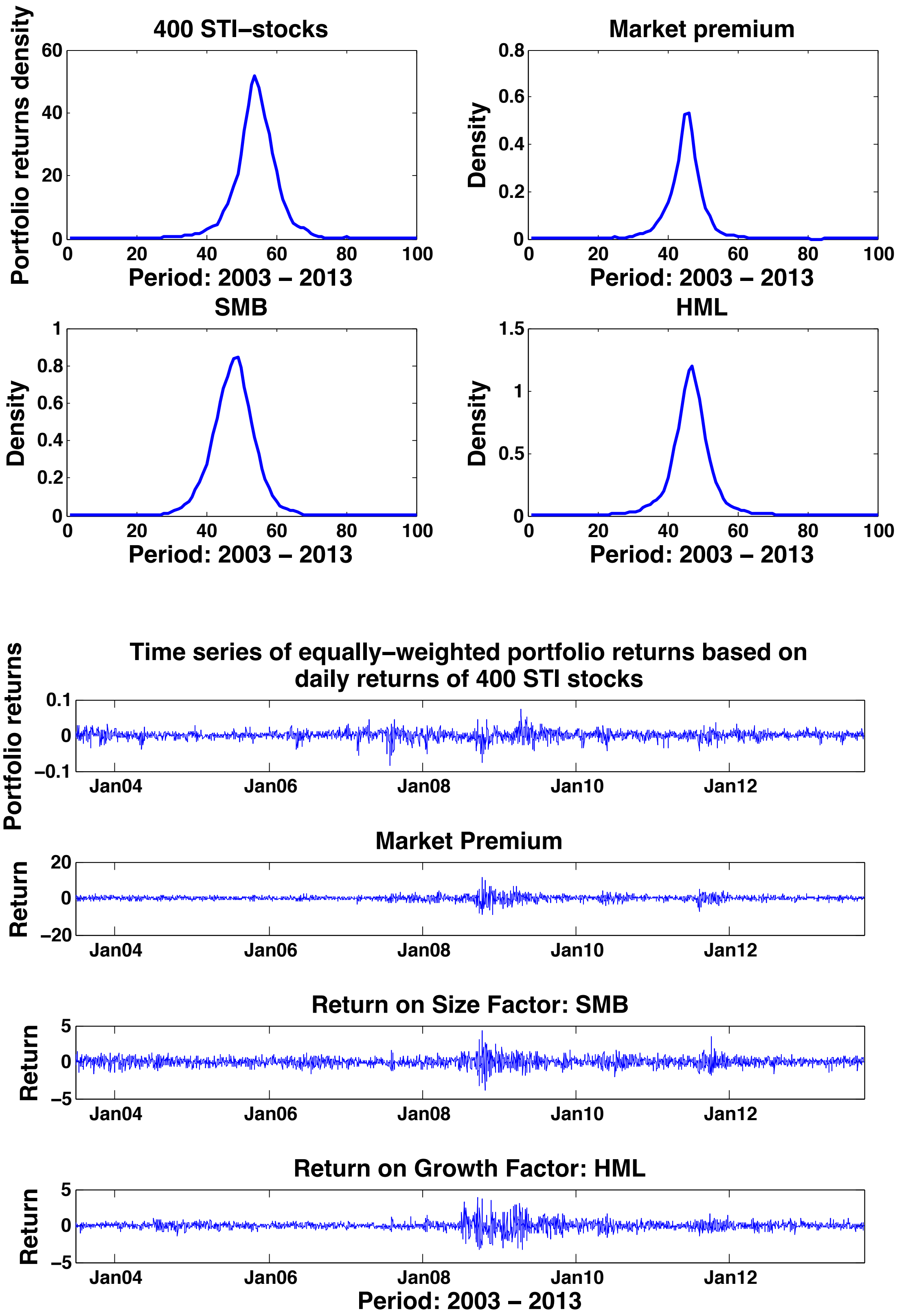


Figure 2: Kernel densities and time series of the equally-weighted 400 STI stocks portfolio and of risk factors.

	Stocks	MP	SMB	HML
Mean	0	0.036	0.0109	0.0164
Min	-0.0836	-8.96	-3.34	-3.79
Max	0.0721	11.35	3.95	4.3
StD	0.0113	1.2663	0.5602	0.5632
Skewness	-0.6188	-0.1378	12.15	-0.0488
Kurtosis	8.5868	13.0583	11.7254	7.7396
JB test	reject	reject	reject	reject
ADF test	reject	reject	reject	reject
KPSS test	reject	do not reject	do not reject	do not reject

Table 1: Data summary statistics.

2.3.3 Empirical set-up

In the empirical set-up the management of a portfolio is mimicked: the covariance matrix is estimated based on the data from $(t-100)$ to t, then the portfolio is formed on the day $(t+1)$ and held for 28 days, then the whole procedure is repeated. It is further assumed that the return on the p-dimensional portfolio is the linear function of the its components:

$$\Delta\Pi = w_1 \Delta X_1 + \cdots + w_p \Delta X_p \tag{25}$$

The portfolios calculated with help of different estimators are then compared based on their out-of-sample standard deviation, average portfolio return and Sharpe ratio.

The choice of the length of the in-sample period is, first of all, a rule-of-thumb and is motivated by the relevance of the information contained in the last 100 days for the portfolio to be formed on day (100+1). Both in-sample estimation period and out-of-sample holding period are made short to possibly account for non-stationarity of the data. A more advanced approach would be to estimated the change points when the regime change occurs and rebalance the portfolio only on those switching points.

In order to provide more robust results, portfolios are formed with $c = 2$ and $c = 4$, i.e. based on 200 and 400 stocks. Since the goal is to analyze the performance of the covariance matrix estimator, only the GMV portfolio is considered, i.e. Markowitz minimization problem is solved with only one constraint requiring that the weights should add up to one. For the same reasonit is sensible to abstract from other possible modifications and extensions such as short-selling and more complex optimization functions which take into account, for example, risk-aversion and other parameters.

2.3.4 Estimation results

In the following section the estimation results for 400 STI stocks-portfolio are presented. The graphical results for 200-stocks portfolio are presented in the Appendix.

	GMV 400, c=4			GMV 200, c=2		
	Return	StD	Sharpe	Return	StD	Sharpe
Shrink to Constant corr	-	-	-	-	-	-
Shrink to Identity	0.0072	3.4958	0.0001	0.0045	3.4071	-0.0016
Shrink to Diagonal	0.0096	7.7508	-0.0025	0.0104	8.0986	-0.0011
Shrink to Market	0.0091	7.7857	-0.0018	0.0125	8.1809	0
Shrink to Two param	0.0076	3.4915	0.0002	0.0046	3.4072	-0.0016
FFL	0.0026	7.3839	-0.0027	0.0040	7.6280	-0.0036
SIM	-0.0013	7.2687	-0.0022	0	7.4363	-0.0031

Table 2: Annualized standard deviations, average annualized returns and Sharpe ratios for 400- and 200-stocks portfolios with different covariance estimators, in %, STI, 2003-2013. Note: Sharpe ratios are not annualized.

The performance of different covariance matrix estimates can be analyzed based on the averages of out-of-sample standard deviations and returns which are summarized in Table 2. First of all, it requires to clarify which quantities are reported here. Annualized standard deviation is measured in the following way: for each of 95 samples consisting of 28 days the standard deviation of the portfolio is calculated. Then the standard deviations are annualized by multiplication with the factor $\sqrt{12}$. The returns are computed for the whole time series and then averaged and annualized. The Sharpe ratios are calculated for the whole time-series and averaged. Sharpe ratios are calculated since considering only standard deviations may be misguiding. Moreover, the Sharpe ratio is essentially the market price of risk: it shows how the investor is compensated for the additional risk taken, therefore, it can help to assess the portfolio performance better.

One observes that the minimum variance is obtained for both GMV 400- and 200-stocks portfolio when covariance matrix estimation based on shrinkage to identity and to a two-parameter matrix are used. The difference in standard deviations with other four competing estimators is rather large. Moreover, only with these two estimators one obtains positive Sharpe ratios. The outcomes for 200-stocks portfolio unfortunately do not back-up these results. However, the ordering of different covariance matrix estimators is still preserved. Analysis of the Fig. 3 where the Sharpe ratios for 95 samples are presented does not allow to make further conclusions as it turns out to be rather volatile.

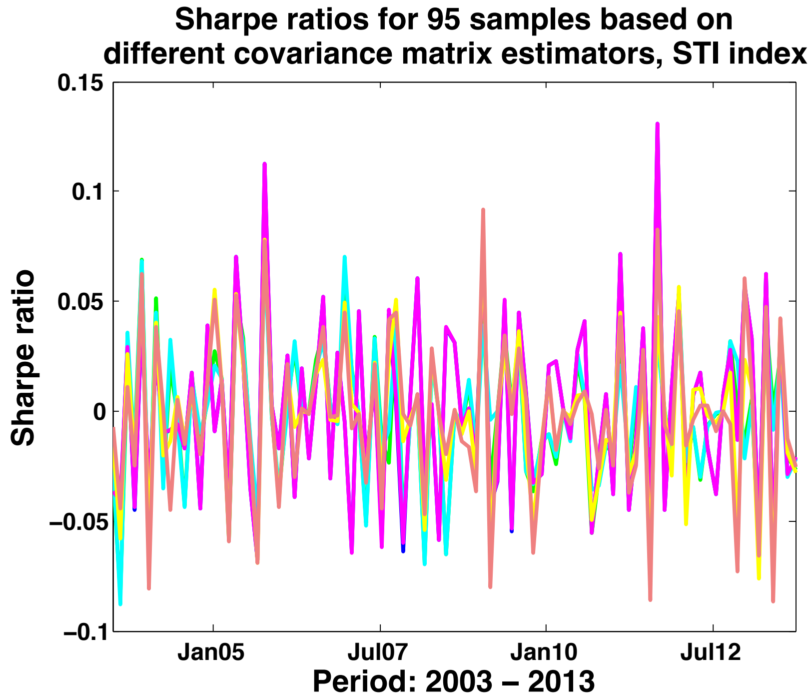


Figure 3: Sharpe ratios for 400 STI-stocks GMV portfolio with different covariance estimators; blue: shrinkage to identity, green: shrinkage to diagonal m., cyan: shrinkage to market, magenta: shrinkage to two parameteres, yellow: FFL, coral: SIM.

Investigation of visual representation of the results provides more intuition on the character of the different estimation techniques. For example, on the Fig. 4 where average standard deviations for GMV 400-stocks portfolio are presented one can easily identify two groups of estimators: coral, cyan, yellow and green correspond to SIM estimator, shrinkage to the market, FFL estimator and shrinkage to a diagonal matrix, respectively. Their lines almost coincide on the graph. Two other lines, magenta and blue, which correspond to shrinkage to two parameters and shrinkage to identity, respectively, form another group and provide much lower standard deviation on average. The results for a GMV 200-stocks portfolio (see Appendix) display a similar pattern - the shrinkage to identity and shrinkage to two parameter gives somewhat 'dumpened' line of standard deviations in comparison to the other four estimators.

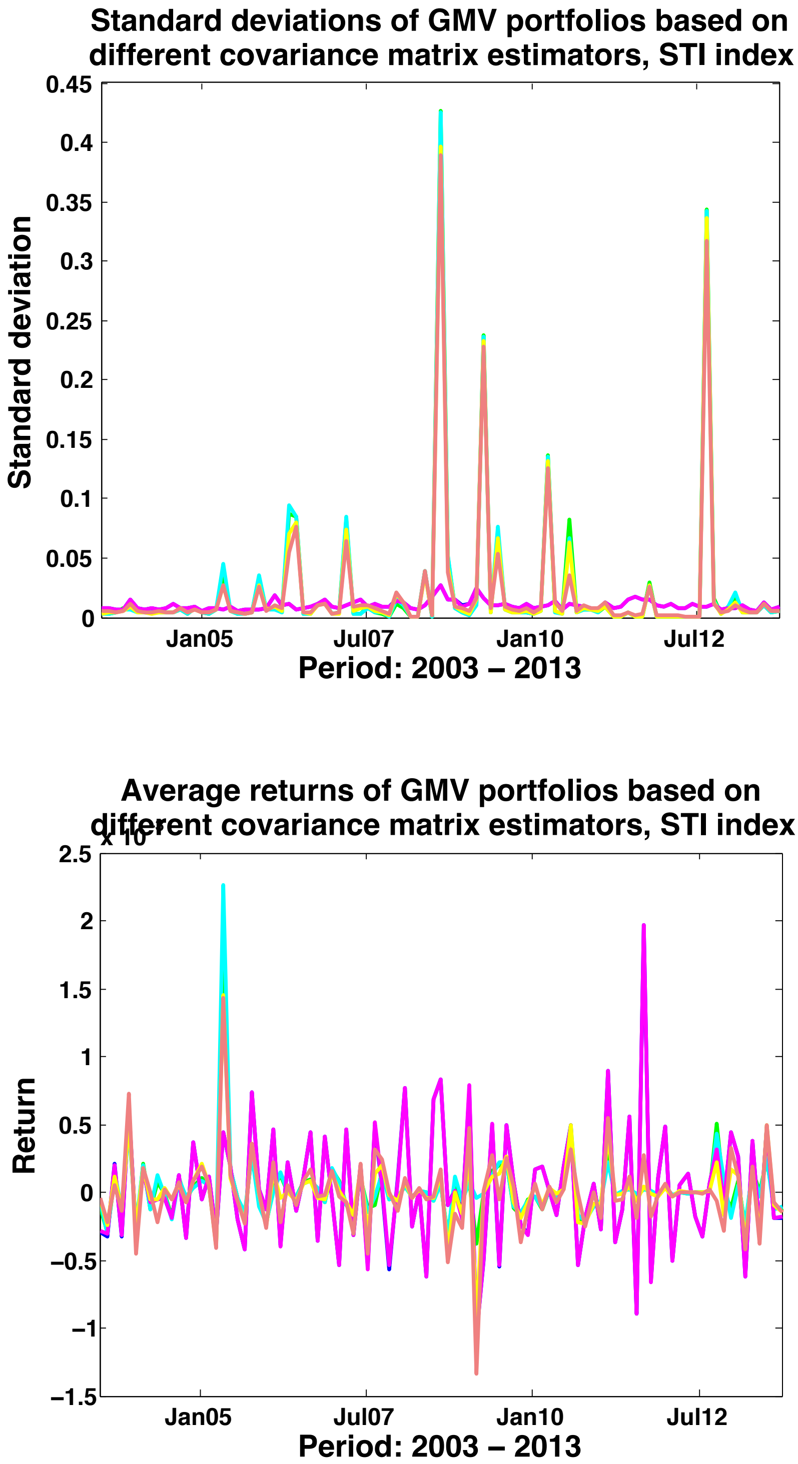


Figure 4: Standard deviations and average returns of 400 STI-stocks GMV portfolio with different covariance estimators; blue: shrinkage to identity, green: shrinkage to diagonal m., cyan: shrinkage to market, magenta: shrinkage to two parameteres, yellow: FFL, coral: SIM.

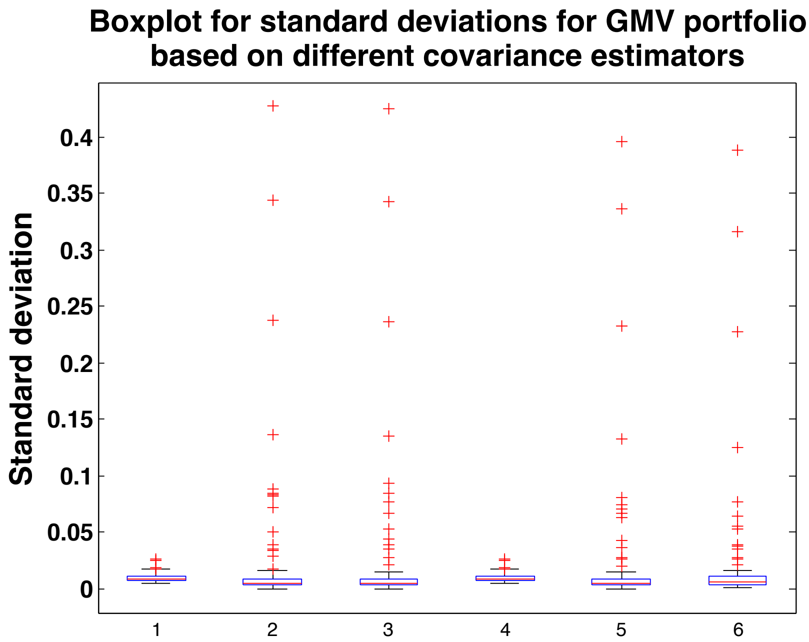
One possible explanation for this distinguishable differences between estimators would be that both shrinking to identity and to a two parameter covariance matrix assume that the variances are the same and covariances are either zero, former case, or the same, latter case. In this sense these estimators are 'tempering' the large deviations of the covariance matrix and force them to stay constant by means of shrinking, whereas other estimators such shrinkage to the market, to a diagonal matrix as well as FFL and SIM estimator are either accentuating the differences in variances over time or even (possibly) amplify them.

To support this observation it is interesting to note that the spikes on the graph of standard deviations occur around 2008, 2009 2010 and late 2012. The first spike occurs around the time when Singapor was hit by the global economic crisis after July 2008. Thus, the increase in stock market volatilities are reflected in the first spike corresponding to the four estimators mentioned above. The second lower spike around 2009 also corresponds to an increase in volatilites financial crisis of 2009. Similarly, other spikes aslo can be attributed to an increase in stock market volatilities.

Interestingly, **Ledot and Wolff (2003)** perfofrm the similar empirical test of various covariance matrix estimators and arrive at the conclusion that the 'best' performing ones (in the sense of providing the lowest standard deviation) are the shrinkage to identity and shrinkage to the market. These results hold partially true also for the conducted empirical study, however, in case of **Ledoi and Wolff (2003)** unfortunately no explanation for this phenomena was provided.

Moreover, the shrinkage estimators, i.e. shrinkage to the market and to a diagonal matrix, have almost the same trajectory as the estimators based on the imposition of the structure. This is remarkable as it is often argued that the advantage of the shrinkage estimators lies in the fact that it exploits the stock data itself and does not require customization of the factors and other information in order to reveal the covariance structure.

Another demonstration of above mentioned 'volatility-tempering' feature of the two shrinkage-based estimators (magenta and blue) is even more pronounced in boxplot representation of the results. Note: There are 95 standard deviations and 2660 returns are plotted.



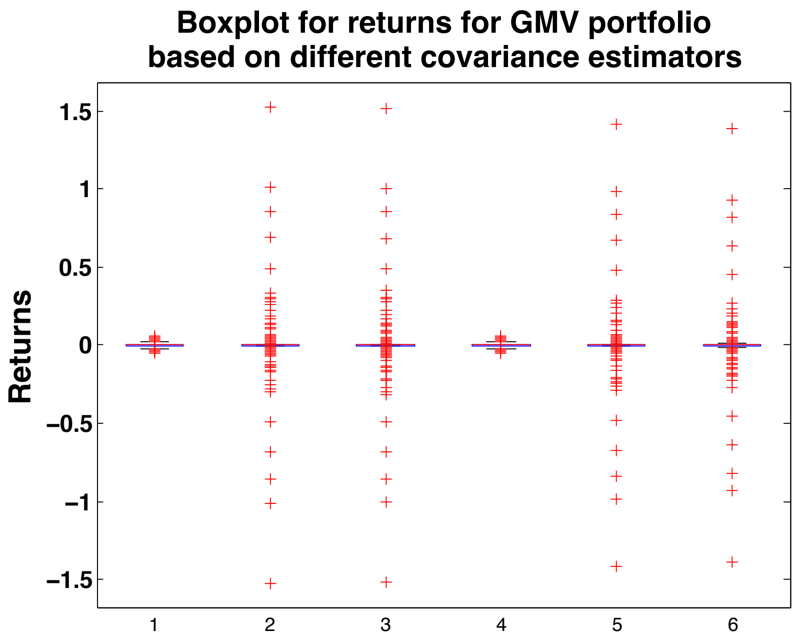


Figure 5: Boxplot for monthly standard deviations and returns of a 400 STI-stocks GMV portfolio with different covariance estimators, 1: shrinkage to identity, 2: shrinkage to diagonal m., 3: shrinkage to market, 4: shrinkage to two parameteres, 5: FFL, 6: SIM.

Another important issue to look at is the evolution of the shrinkage intensities for different estimators. First of all, the shrikage intensities are not stable and are quite volatile over time. Moreover, as one sees on the Fig. 6 the shrinkage intensities can be separated into two groups: cyan and green corresponding to the shrinkage to the market and to a diagonal matrix, respectively, and magenty and blue corresponding to shrinkage to a two parameter matrix and to an identity matrix, respectively. (The red line corresponds to a shrinkage intensity for a constant correlation model and this model did not perform well returning the shrinkage intensity equal to 1. Thus, this covariance matrix estimator is omitted from the further analysis. For a 200-stocks portfolio this estimator performs well, however, only for the first part of the sampel period, i.e. till around 2009.) One observes that although all lines move to the same direction, shrinking to a market and to a diagonal matrix is more amplified than shrinkage to idenity and to two parameter matrix. Intuitively, it is clear that the shrinakge intensity should be higher when less structure on the shrinkage target is imposed.

Moreover, the optimal shrinkage intensity depends on the correlation between estimation error on the sample covariance matrix and on the shrinkage target. If the estimation error on the sample covariance matrix and on the shrinkage target are positively correlated, the advantage of combining the information is diminishing. (See **Ledoit and Wolff (2000)**). In other words, when the shrinkage intensity is, e.g., 80 % this means that "there is four times as much estimation error in the sample covariance matrix as there is bias in the [shrinkage target]". Logically, low shrinkage intensity can be a result of a large bias of the shrinkage target as well as little estimation error of the sample covariance matrix. Since the second case is unlikely as it is known that if $p > n$ the eigenvalues of the sample covariance matrix are distorted, it is presumably the case that the structure imposed on the data contains too large bias. This is intuitively true, since the stocks in crises or booms are known to be highly correlated - the feature which is ignored when the shrinkage target is either an idenity matrix or a matrix with equal variances and correlations. This feature can possibly be exploited by the investor who prefers to keep the standard deviation of the portfolio at the lowest level during the crises.

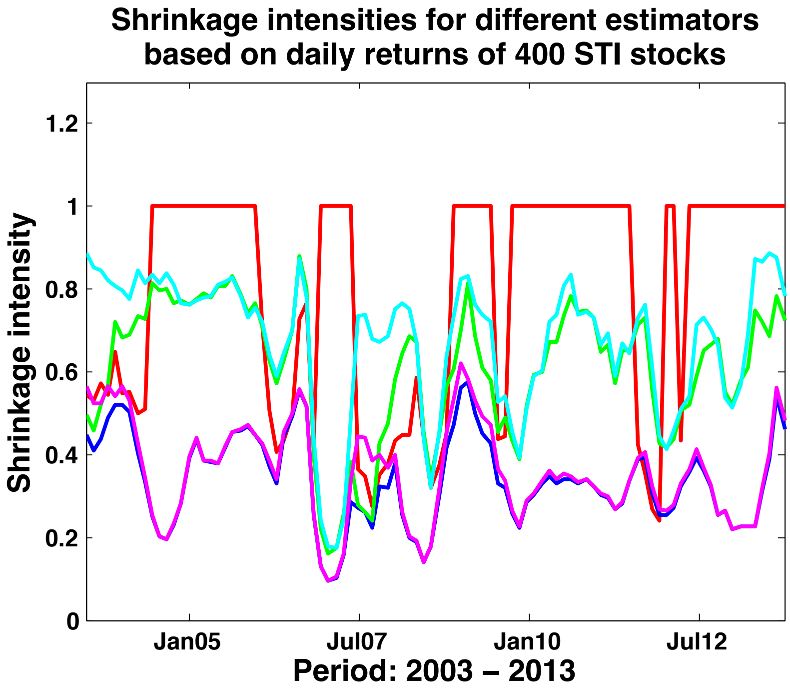


Figure 6: Shrinkage intensities for different covariance estimators (400 stocks); red: shrinkage to constant correlation, blue: shrinkage to identity, green: shrinkage to diagonal m., cyan: shrinkage to market, magenta: shrinkage to two parameteres.

2.4 Comparison of estimation methods

Emprical comparison of two competing approaches to covariance matrix estimation allows to say that the critique of the factor-based models is not justified in the sense that factor structure can be biased since one could observe that both estimators based on factor structure and on shrinakge principle lead to the similar results. What is more important to note is that it is not the approach itself, i.e. factor-based estimation vs shrinkage, which delivers a different result, but it is the structure imposed on the covariance which plays a decisive role. In other words, the difference comes with the amount of dimension reduction imposed on the estimator.

Particularly, it is interesting to see that shrinkage to a diagonal matrix almost coincides with factor-based estimators. Since in diagonal shrinkage the off-diagonal elements are shrinked and on-diagonal elements stay the same whereas in factor-based model all elements receive the same weight, this means that the variances of the assets are more important than the covariances for the overall portfolio variance.

This conclusion calls for a diversification of the high-dimensional matrix estimation - instead of using merely mathematical assumptions, one can use the information contained in the data, to obtain a more precise estimator. For example, if one knows or can predict that the covariances do not play important role in a certain period or, alternatively, the goal is to restrict the influence of covariances, it is sensible to shrink to an identity matrix or to a diagonal matrix. The same logic can be applied when the correlation structure between the assets is known, i.e. if one knows the range of correlation coefficients and the sign or possibly also the magnitude of such correlation, also multitarget shrinkages can be constructed.

Similar effect can be achieved using the factor-based models if this approach is combined with some kind of thresholding or sparsitency assumption. However, in general, factor-based models will remain more responsive to the changes in variances and it is easier to perform a stringent restriction with shrinkage estimators.

3 Estimation of the Value-at-Risk for a High-Dimensional Portfolio

Modeling risk can be differentiated along two major dimensions: the value to be modeled, for example, value-at-risk, expected shortfall or the entire density of distribution, and the modeling approach: parametric, non-parametric and semi-parametric. In this thesis examples of all the three approaches are presented for estimating the value-at-risk of a portfolio which is rebalanced on a daily basis.

For a given portfolio, probability and a time horizon, value-at-risk is defined as a threshold value such that the probability that the loss on the portfolio over the given time horizon exceed this value is the given probability level (See **Li et al. (2011)**). Technically, VaR is defined as

$$\Pr(\Delta\Pi_t < VaR) = \alpha \tag{26}$$

where $\Delta\Pi_t$ is the change in a portfolio's value from $(t-1)$ to t.

The major difficulty in calculating the VaR lies in the fact that the true distribution is never known. Since the maajor interest lies in analysis of the tails of distribution it is helpful to represent the data visually to get intuition about the nature of the underlying process.

On Fig. 7 the QQ plots of the returns on 400 STI stock portfolios rebalanced on a daily basis are presented. The QQ plot is particularly helpful in revealing the leptokurtic tails of a distribution since the empirical quantiles are plotted against the theoretical quantiles of a normal distibution. If the underlying distribution is normal, the QQ plot should clearly be a 45 % line.

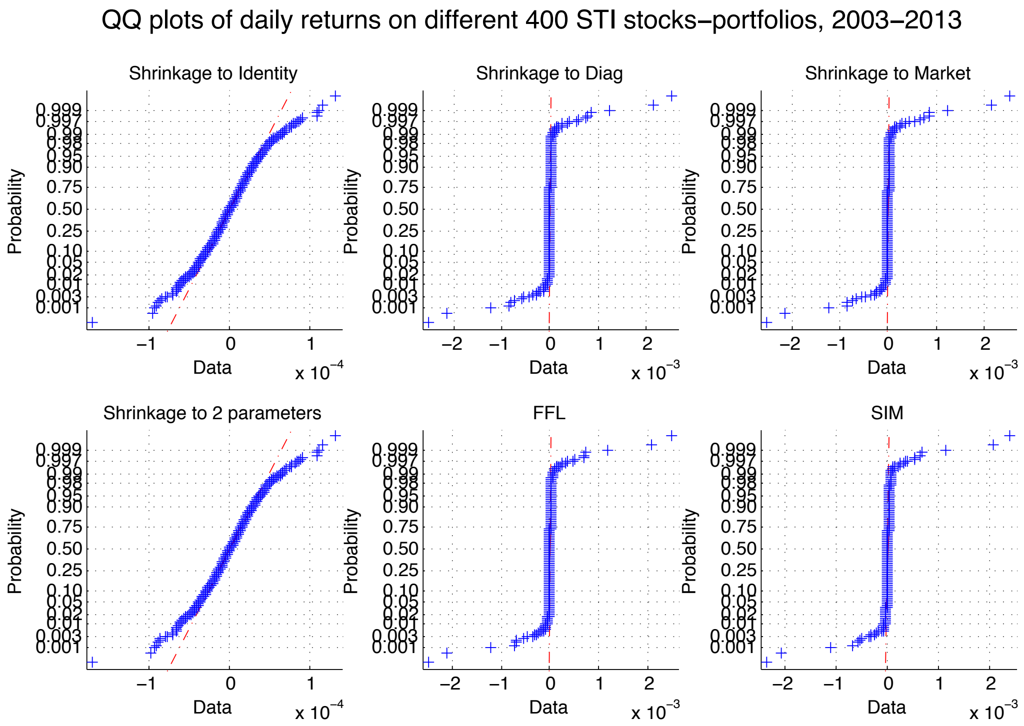


Figure 7: QQ plots for 400 STI stocks-portfolio returns. Note: Portofolios are rebalanced every day, the moving window size is $n = 100$.

Clearly, the underlying distibutions are heavy-tailed. More specifically, the middle parts of the QQ plot coincide with the theoretical quantiles, however, the tails' parts severely deviate. In order to solve this problem, one can employ a certain skewed distribution which will capture the heavy-tails of the empirical distibution, or one can estimate the VaR non- or semiparametrically when there are either no distibutional assumptions or they are imposed only partially.

It is interesting to note that for two covariance matrix estimates, namely, shrinkage to identity and shrinkage to two parameter matrix, the tails are visibly thinner than for other estimates. This might be related to the phenomenon of 'tempered' volatility observed in the previous section. Therefore, it is plausible to deduce that the volatiltiy and kurtosis might be related.

In this section a novel approach to calculating VaR is intoduced, namely, a VaR measure based on the semi-parametric estimation of the density of the returns. For completeness the main methods for VaR calculations are presented and tested empirically. The results for a 400-STI stocks-portfolio formed based on shrinkage to identity are presented graphically, for all other portfolios only the number of exceedances are given.

3.1 Nonparametric VaR

Nonparametric VaR, or historical VaR, is the simplest method to calculate value-at-risk. The procedure is the following: one observes the returns on a portfoilo from day 1 to day 28 from the sample 1, then it is assumed that the empirical quantile of the sample 1 is a good proxy for the VaR for the sample 2, and so on. In other words, the past is the best prediction for the future.

$$\mathrm{VaR}_{t+1}^{\alpha} = q_t^{\alpha}, \text{where t} = 1,\ldots,\text{T, and } \alpha \text{ is the level of quantile.}$$

This approach might be justfied when the window size is small enough, however, it does not protect an investor from a drastical changes such as, e.g., sudden stock market crash or other unexpected events. A refinement to this method is a bootstrap historical simulation approach. The bootsrap draws a sample from a data set (in this case, from a portfolio profit and losses) and obtains VaR, or a specified empirical quantlie, from that sample. This procedure is repeated from 1000 to 10 000 times. The 'best' VaR estimated from the data set is the average of all bootstrapped VaR.

This approach with bootstrapping repeated 1000 times is applied to the portfolios formed by using different covariance matrix estimates. Note: VaR is calcuted for the daily frequency - it is assumed that a portfolio is rebalanced every day.

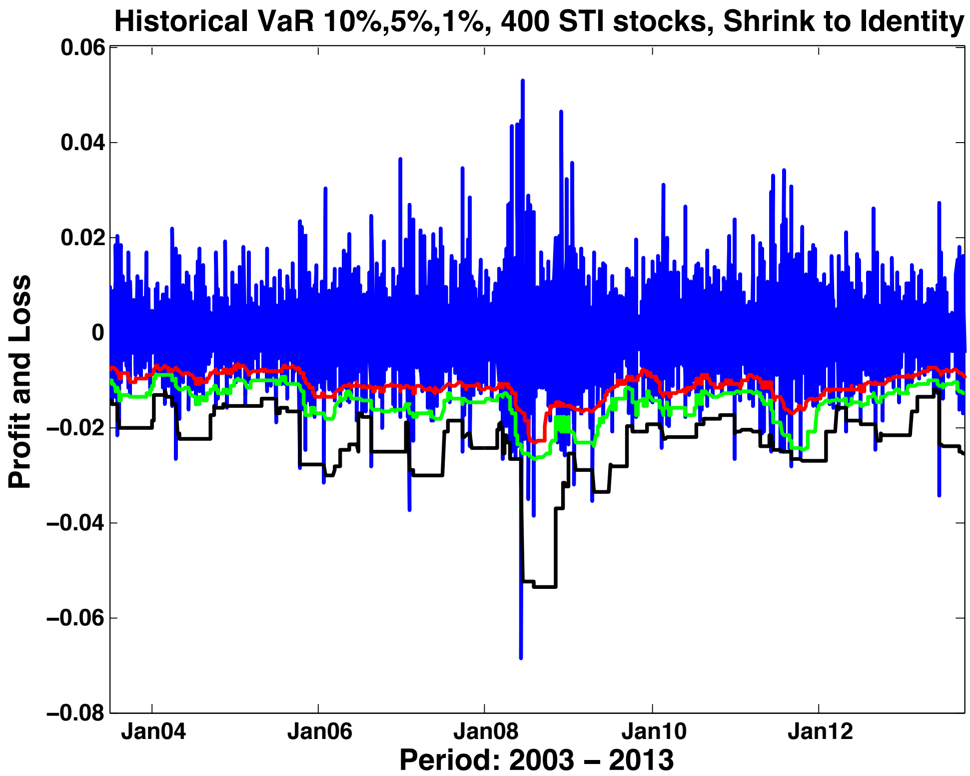


Figure 8: Historical VaR for 400-stocks portfolio based on shrinkage to Identity; red - 10% VaR, green - 5% VaR and black - 1% VaR.

	10%	**5%**	**1%**
Shrink to Identity	252	138	36
Shrink to Diagonal	256	134	21
Shrink to Market	254	134	20
Shrink to Two param	249	140	37
FFL	260	138	22
SIM	247	139	23

Table 3: Number of exceedances for historical VaR for 400-stocks portfolios based on different covariance matrix estimators.

3.2 Parametric VaR

3.2.1 Delta-Normal VaR

Along with historical VaR, delta-normal approach is one of the simplest method to calcuate VaR. It assumes that the portfolio profit and losses are linear and the risk factors are joinly linear distributed. Thus, the portfolio standard deviation can be calculated by using covariance matrix and weights.

$$\Sigma_{portfolio} = w^T \Sigma w \tag{27}$$

Consequently, VaR is defined as follows:

$$\mathrm{VaR}_\alpha = z_\alpha \sqrt{w^T \Sigma w} \tag{28}$$

where z_α is the standard normal quantile. Note that here the mean of a portfolio is assumed to be close to zero.

The main advantage in using the delta-normal method lies in its simplicity, however, the assumption of normal distribution often leads to underestimation of the extreme outcomes. The disadvantage lies in a non-realistic distibutional assumption imposed on a portfolio distibution. This is also proven empirically - as one can see the number of exceedances are too high to employ this methodology in practice.

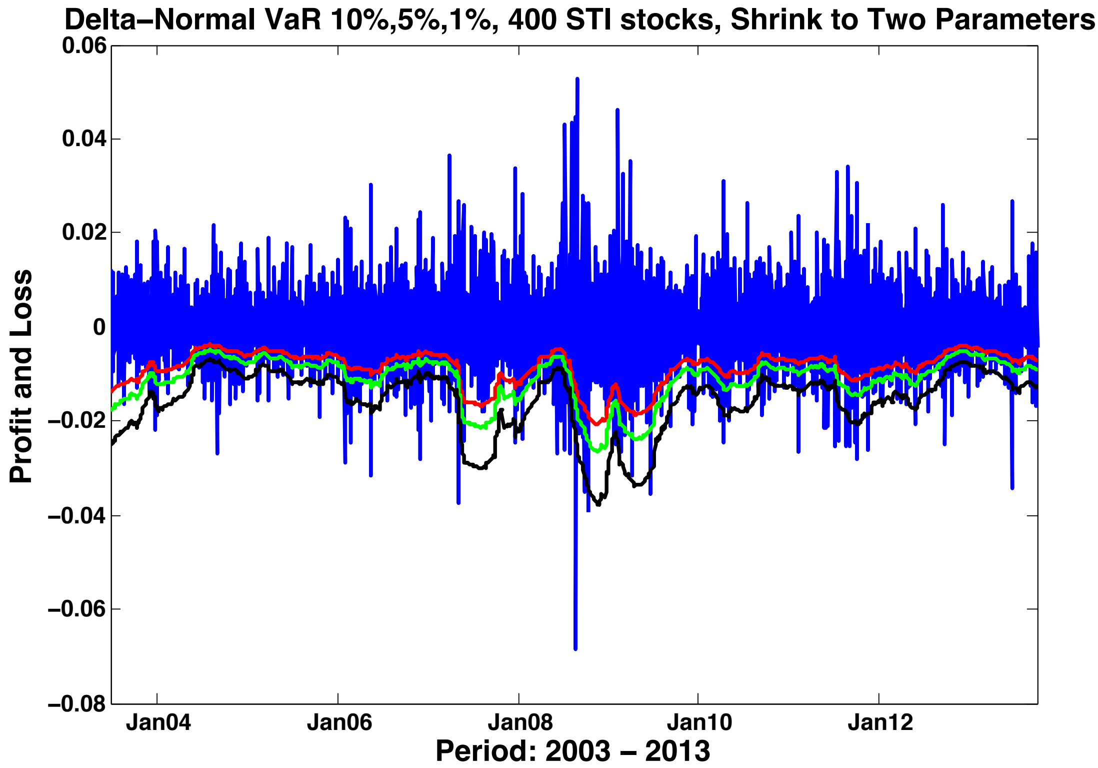


Figure 9: Delta-normal VaR for 400-stocks portfolio based on shrinkage to Identity; red - 10% VaR, green - 5% VaR and black - 1% VaR.

	10%	**5%**	**1%**
Shrink to Identity	492	366	198
Shrink to Diagonal	401	281	115
Shrink to Market	405	290	123
Shrink to Two param	225	116	64
FFL	405	323	174
SIM	585	420	191

Table 4: Number of exceedances for delta-normal VaR for 400-stocks portfolios based on different covariance matrix estimators.

3.2.2 Monte-Carlo VaR

Monte-Carlo VaR is similar to historical simulation method with the exception that the distribution assumption about the stochastic process is made. After obtaining the estimates for the expected return and covariance matrix, one simulates, for example, from a normal or t-distribution from 1 000 to 10 000 samples with given expected return and covariance and calculates the α-quantile for each sample. Clearly, the more frequently the estimates for the expected return and the covariance matrix are updated, the more precise are the estimated VaR values. The 'best' VaR estimated from the data set is the average of all quantiles of interest from simulated samples.

The disadvantages of this method are that it can be time-consuming and hard to implement due to limited hardware capacities and that the distibution assumptions may be wrong and do not take into account the extreme values which are of primary interest in case of VaR calculations.

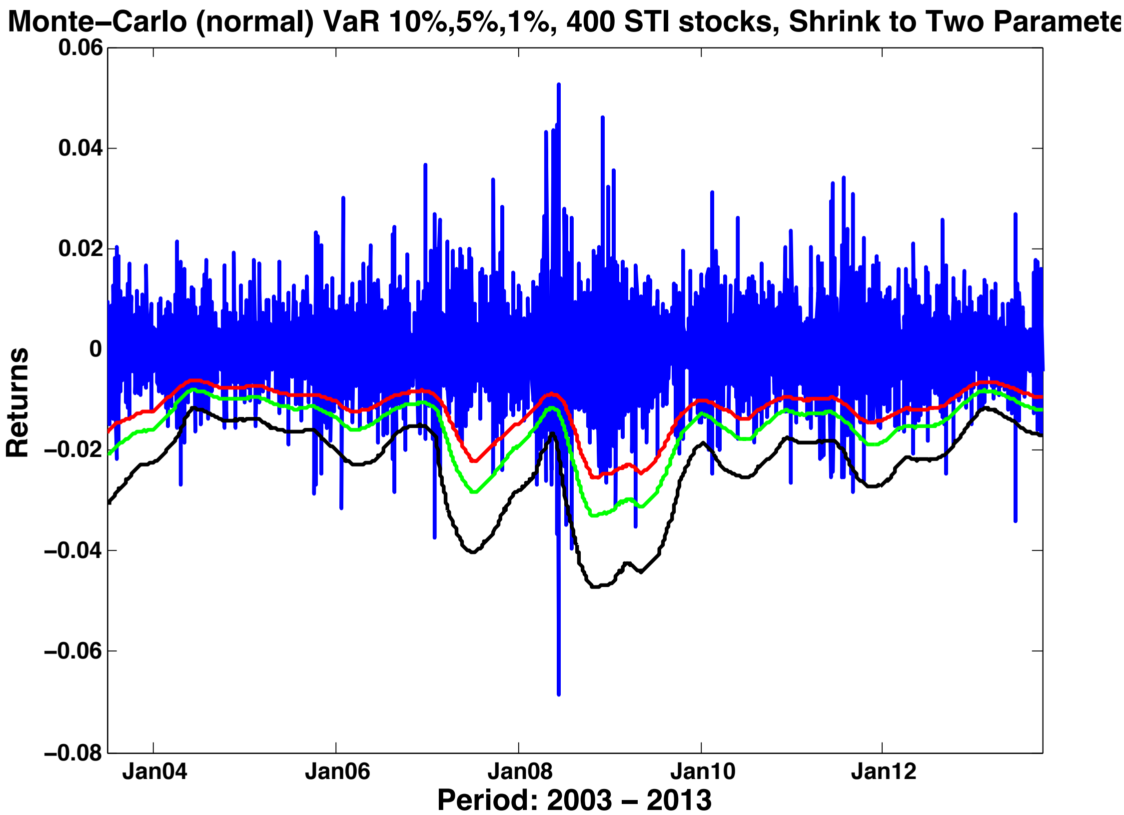


Figure 10: Monte-Carlo VaR for 400-stocks portfolio based on shrinkage to Identity; red - 10% VaR, green - 5% VaR and black - 1% VaR.

	10%	**5%**	**1%**
Shrink to Identity	501	270	178
Shrink to Diagonal	384	282	202
Shrink to Market	421	308	234
Shrink to Two param	347	144	52
FFL	509	303	261
SIM	510	312	151

Table 5: Number of exceedances for Monte-Carlo VaR for 400-stocks portfolios based on different covariance matrix estimators.

3.3 Semiparametric VaR

In this subsection the novel method for calculating the VaR is presented. It is based on the semparametric estimation of multivariate generalized elliptical density. The methodology is developed in **Fan, Härdle and Okhrin (2012)**. The core of the method lies in nonparametric estimation of a projection of multivariate density. This method besides being simple and intuituve in implementation also allows to account for the heavy-tails of the distibution explicitely. Moreover, due to its partial nonparametric nature it captures skewness in returns' distributions without explicit introduction of a skewness parameter or a function. In the following subsections, first the theoretical outline is presented, then the empirical results are analyzed. The theoretical background on generalized elliptical distributions is based on **Fang et al. (1990)**, **Branco and Dey (2001)** and **Hult and Lindskog (2002)**.

3.3.1 Theoretical outline

Elliptical distribution can be thought of as generalization of normal distribution. It were introduced by **Kelker (1970)** and further investigated by **Cambanis, Huang, and Simons (1981)** and by **Fang, Kotz, and Ng (1990)**. A well-written overview on generalized elliptical distributions is provided by **Frahm (2004)**. Definitions and theorems provided below can be found in **Frahm (2004)**.

Theorem 1: *A random variable* $Y \sim El_p(\mu, \Sigma, \varphi)$ *with* $r(\Sigma) = k$ *is said to be* *generalized elliptically distributed if and only if*

$$Y = \mu + R\Lambda U^{(k)} \tag{29}$$

where $U^{(k)}$ is a k-dimensional random vector uniformly distributed on S^{k-1}, R is a non-negative random variable being stochastically independent of $U^{(k)}$, $\mu \in R^p$ and $\Lambda \in R^{p\times k}$.

Here, R determines the shape of the distribution, in particular, tails, and μ determines the location of the random vector Y (see Frahm (2004)).

The density function of a multivariate elliptical variable is defined as follows.

Theorem 2: *Let $Y \sim El_p(\mu, \Sigma, \varphi)$ where $\mu \in R^{d\times d}$ and $\Sigma \in R^{p\times p}$ is positive semidefinite with $r(\Sigma) = k$. Then Y can be represented stochastically by $Y = \mu + R\Lambda U^{(k)}$ with $\Lambda\Lambda^T = \Sigma$ according to Theorem 1. Further, let the cdf if R be absolutely continuous and S_Λ be the linear subspace of R^p spanned by Λ. Then the pdf of Y is given by*

$$y \to f_y(y) = |det(\Lambda)|^{-1} g_R\{(y-\mu)^T\Sigma^{-1}(y-\mu)\} \tag{30}$$

where $x \in S_\Lambda \setminus \{\mu\}$ and

$$t \to g_R(r) := \frac{\Gamma(k/2)}{2\pi^{k/2}}(\sqrt{t})^{-(k-1)} f_R(\sqrt{t}) \tag{31}$$

where $t > 0$ and f_R is the pdf of R.

In order to be able to estimate this density one can rearrange the formula above to obtain the generator function which depends on r:

$$g(r) = \frac{\Gamma(p/2)}{2\pi^{p/2}} r^{1-p/2} g_R^2(r)$$

This reformulation allows to separate estimation procedure into two parts: $g_R^2(r)$ can be estimated non-parametrically, then the estimate of $g(r)$ can be obtained.

3.3.2 Estimation procedure

The exact procedure is the following:

1. Estimate mean and covariance matrix. Here, the estimator of the mean is assumed to be the mean of the last 100 observations. The estimators of covariance matrices are taken from the first part of this thesis.

2. Estimate kernel density of transformed variables:

$$\hat{k}(x, h, \hat{\Sigma}) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - \hat{r}}{h}) + K(\frac{x + \hat{r}}{h})$$

   where $\hat{r} = \{(Y - \hat{\mu})^T \hat{\Sigma}^{-1} (Y - \hat{\mu})\}$, h is the bandwidth for the kernel density estimation. Here, the Silverman's tule for optimal bandwidth calculation is used, i.e. $\hat{h} = 1.06\sqrt{Var(\hat{r})n^{-1/5}}$, and as a kernel a Gaussian kernel is used which is defined as $K(u) = \frac{1}{\sqrt{2\pi}} exp(-\frac{u^2}{2})$.

3. Obtain the estimator of $g(r)$ as follows:

$$\hat{g}(r) = \frac{\Gamma(p/2)}{2\pi^{p/2}} r^{1-p/2} \hat{k}(x, h, \hat{\Sigma})$$

4. Obtain the estimator of multivariate generalized elliptical density:

$$\hat{f}(y|\hat{\mu}, \hat{\Sigma}) = |\hat{\Sigma}|^{-1/2} g^{(p)}\{(Y - \hat{\mu})^T \hat{\Sigma}^{-1} (Y - \hat{\mu})\}$$

On the Fig. 11 (top) one can see the density $g_R^2(r)$ and $log(g(r))$ estimated for stock returns with different covariance matrix estimators. On the Fig. 11 (bottom) the evolution of the density is plotted with the covariance matrix estimator based on shrinkage to identity. The blue color denotes the 'calm' period (450 samples from 2003 with a moving window $n = 100$) and the red color denotes 'crisis' period (800 samples from 2008 and 2009 with a moving window of $n = 100$). One can see that the increased volatility in the crisis period is reflected in more dispersion around the location (see graph of $g_R^2(r)$) and in fatter tails (see graph of $log(g(r))$).

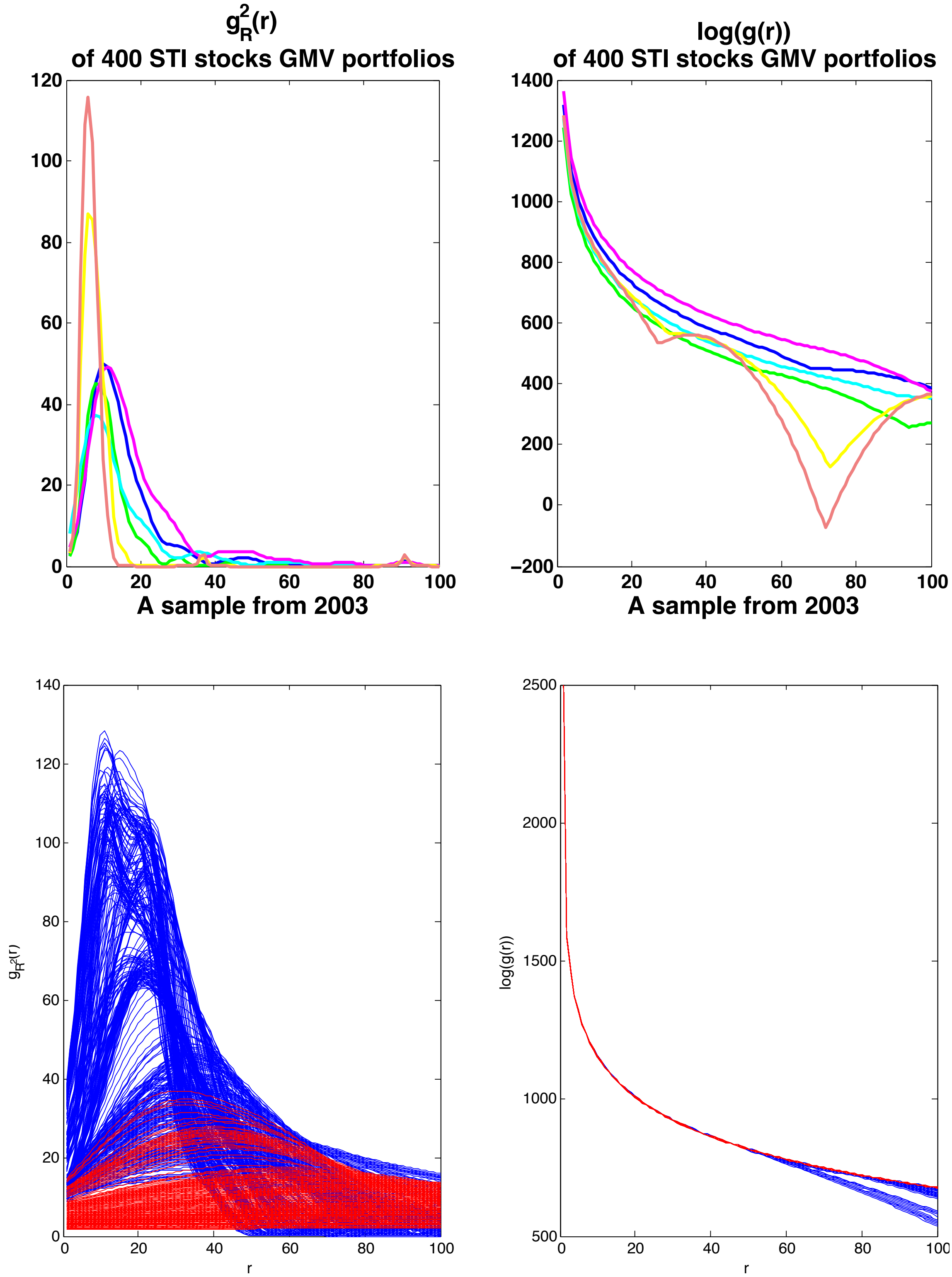


Figure 11: Top: $g_R^2(r)$ and $log(g(r))$ estimated for a sample of STI stock returns in 2003, $n = 100$, $p = 400$;400 STI-stocks GMV portfolio with different covariance estimators; blue: shrinkage to identity, green: shrinkage to diagonal m., cyan: shrinkage to market, magenta: shrinkage to two parameteres, yellow: FFL, coral: SIM; bottom: evolution of $g_R^2(r)$ and $log(g(r))$ with shrinkage to identity estimator in 'calm' period 2003 (blue) and 'crisis' period 2008-2009 (red).

3.3.3 VaR for elliptical distributions

If the pricing function of a portfolio is linear in its risk factors, i. e. if the return on a portfolio is a weighted sum of the returns on its components, then the VaR of a portfolio is defined as follows:

$$Prob\{\Delta\Pi(t) < VaR_\alpha\} = \alpha$$

where $\Delta\Pi(t) = w_1Y_1 + w_2Y_2 + \cdots + w_pY_p$ is profit and loos of a portfolio over a time horizon $t+1$ to t with $\Delta\Pi = \Pi(t) - \Pi(0)$ denoting a change in the value of a portfolio.

This general definition of a VaR can be formulated in terms of elliptical distribution as follows:

$$\alpha = |\Sigma|^{-1/2} \int_{wy<-VaR_\alpha} g\{(Y-\mu)^T\Sigma^{-1}(Y-\mu)\}dy$$

In practice, one can obtain the quantile of elliptical distribution by solving the following equation:

$$G(s) = \int_s^\infty K(s,u)g(u)du$$

where the kernel K is given as $K = \frac{\pi^{\frac{n-1}{2}}}{2\Gamma(\frac{n-1}{2})} \int_{\sqrt{s}}^{\sqrt{u}} (u-z_1)^{\frac{n-3}{2}dz_1}$ and $u = r^2 + z_1^2$, $|z|^2 = z_1^2 + |z'|^2$ with $z' \in R^{n-1}$ and $x = (Y-\mu)A^{-1}$ and $x = zR$. For a detailed derivations see Kamdem (2005).

3.3.4 Empirical results

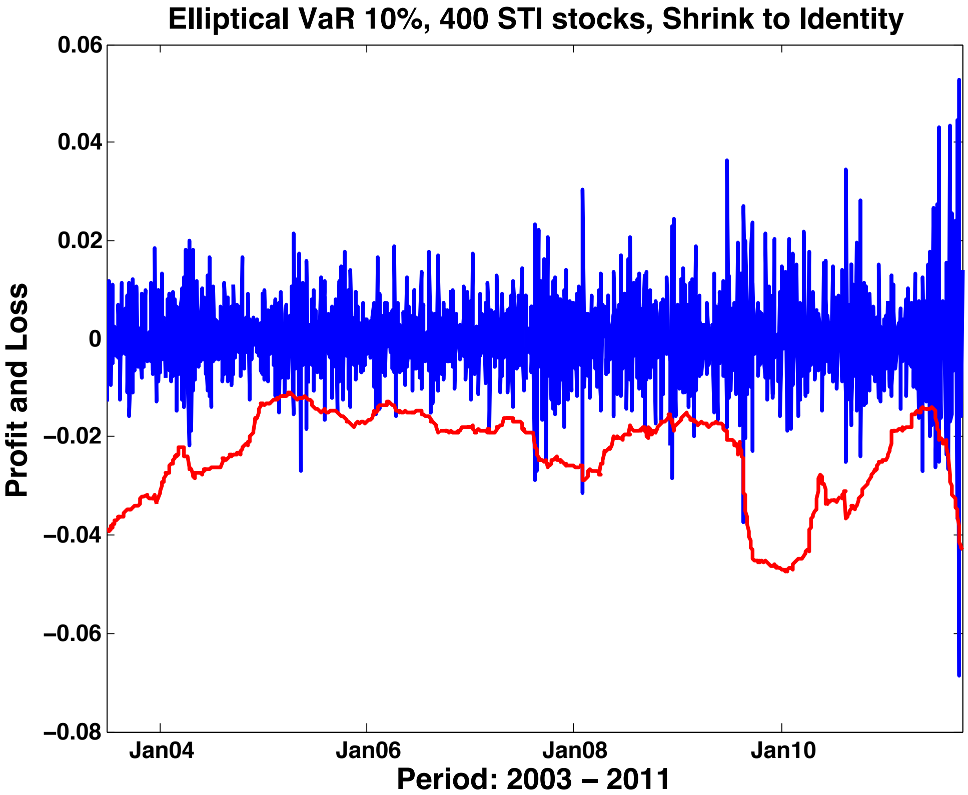


Figure 12: Elliptical VaR for 400-stocks portfolio based on shrinkage to Identity; red - 10% VaR.

	10%
Shrink to Identity	85
Shrink to Diagonal	90
Shrink to Market	84
Shrink to Two param	60
FFL	91
SIM	94

Table 6: Number of exceedances for elliptical VaR for 400-stocks portfolios based on different covariance matrix estimators.

3.4 Comparison of estimation methods

Despite of strikingly different distributional characteristics of portfolio returns the results of the VaR estimation do not differ so drastically. The results for the historical simulation are of similar magnitude for all portfolios. This is an expected result, since the value-at-risk depends only on the values of the returns themselves.

For delta-normal VaR one can see a certain positive bias for the two estimators for which portfolio returns were more Gaussian-distributed. The same holds true for the Monte-Carlo simulations based on normal distribution. However, for other returns series the number of exceedances are drastically high which makes these methods certainly impossible to apply in practice.

This justifies introduction of a new method which takes into account leptokurtic tails of distribution, namely, semi-parametric VaR considered in the last subsection. It should be noted that the numerical procedure for integration and quantile search is quite lengthy, therefore, the estimation period was reduced up to the end of 2011 and only 10

4 Conclusion

In the following thesis several critical issues of the modern finance were addressed: first, several covariance matrix estimators for high-dimensional data were analyzed; secondly, the issue of leptokurtic tails was addressed in value-of-risk computation. This analysis based on empirical data allows to draw certain conclusions with respect to the different methods of high-dimensional covariance matrix estimation as well as with respect to the value-at-risk calculation.

In the theoretical literature there exist three main approaches to covariance matrix estimation when the dimension is greater than the sample size: factor-based approach, shrinkage approach and direct operations on sample covariance matrix. Although based on different methodology all of them aim at obtaining a lower dimensional representation of high-dimensional data. Moreover, some of them despite of being contrasted to each other seem to arrive at the similar result. Thus, for example, shrinkage estimators when the number of parameters to be estimated is high (in this case, p parameters are considered to be a 'high' number to estimate) produce similar portfolio volatility as covariance matrix estimators based on factor models. Therefore, one can conclude that the true distinction is not in the method itself, but in the magnitude with which the high-dimensional data is reduced to a lower-dimensional space - the more the data is 'squeezed' into lower subspace, the greater are the differences between the estimators.

Given that one can certainly think of a situation when common factors (suggested by Fama and French, for example) are indeed not relevant for the market of interest. In such a situation the shrinkage estimators developed by Ledoit and Wolff can serve as substitutes for the factor-based models.

It is important to note that different methods perform differently in calm and crisis periods. Although the maximum shrinkage result in general in lower volatility a portfolio, in certain periods less dimension reduction perform as good sa high amount of shrinkage or there are even periods when estimators based on less dimension reduction perform better. For an investor this means that it can be beneficial to switch between various estimators depending on the expected situation on the market. High amount of shrinkage can be also viewed as a hedge against increased volatility .

Moreover, as a result of optimization with different covariance estimators one obtains portfolios have different values which are distributed differently. Portfolio returns with more shrinkage are more 'normally' distributed than the portfolio returns with less amount of shrinkage or factor-based models. This is explained by the fact that the latter estimators are more responsive to the market, thus, the volatility of the market is reflected in the volatility of a portfolio. Therefore, the methodology of value-at-risk calculations should be adjusted respectively.

For the future research it is interesting to consider different reduction methods which will incorporate the information available in the market. For example, if one knows that certain stocks are positively correlated, then shrinkage to identity can be substituted for the shrinkage to constant variance and certain predefined positive covariances. Moreover, multivariate shrinkage targets should be considered. Valuation of risk should be adjusted for different estimators, otherwise, the value-of-risk can be over- or underestimated.

5 Bibliography

Bai, J. and Ng, S. (2008). Large Dimensional Factor Analysis. *Foundations and Trends in Econometrics 3/2*: 89-163.

Banerjee, A., I. Masten, and M. Massimiliano (2006). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Elsevier.

Bickel, P. J., and Levina, E. (2008a). Covariance regularization by thresholding. *The Annals of Statistics, 36(6)*: 2577-2604.

Bickel, P. J., and Levina, E. (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 199-227.

Branco, M. D., and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis, 79(1)*: 99-113.

Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis, 11(3)*: 368-385.

Efron, B.,, and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association 70/350*: 311-319.

Fama, E. F., and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33(1)*: 3-56.

Fan, J., Fan Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147.1*: 186-197.

Fan, J., Liao Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics 39/6*(2011): 3320-3356.

Fan, J., W. Härdle, and O. Okhrin (2012). Semiparametric Estimation for very highdimensional Elliptical Distributions. Forthcoming.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). Symmetric Multivariate and Related Distributions - Monographs on Statistics and Applied Probability. London: Chapman and Hall Ltd. MR1071174.

Frahm, G. (2004). Generalized elliptical distributions: theory and applications. Diss. Universität zu Köln, 2004.

Frost, P. A., and Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis, 21(3)*: 293-305.

Giannone, D., L. Reichlin, and D. Small (2007). Nowcasting: The real- time informational content of macroeconomic data. *Journal of Monetary Economics*: 45-89.

Golosnoy, V., and Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *The European Journal of Finance, 13(5)*: 441-458.

Higham, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl. 103*: 103–118.

Hult, H., and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied probability, 34(3)*: 587-608.

Kamdem, J. (2005). Value-at-Risk and expected shortfall for linear portfolios with elliptically distributed risk factors. *International Journal of Theoretical and Applied Finance, 8*: 537–551.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya A 32*: 419-430.

Ledoit, O., and Wolf, M. (2003). Honey, I shrunk the sample covariance matrix. *UPF Economics and Business Working Paper*: 691 - 702.

Ledoit, O., and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10(5)*: 603-621.

Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88(2)*: 365-411.

Ledoit, O., and Wolf, M. (2013). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. Available at SSRN 2198287.

Li et al. (2011). Approaches to VaR. Standford University class materials.

Muirhead, J. (1987). Developments in eigenvalue estimation. *Advances in Multivariate Statistical Analysis*: 277-288.

Ledoit, O., and Wolf, M. (2013). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. Available at SSRN 2198287.

Schäfer, J., and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical*

Applications in Genetics and molecular Biology, 4(1): 32.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science, 9(2)*: 277-293.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley symposium on mathematical statistics and probability 1/399*: 197-206.

Stock, J. H., and Watson, M. W. (2007). Why has US inflation become harder to forecast?. *Journal of Money, Credit and Banking, 39(s1)*: 3-33.

6 Appendix

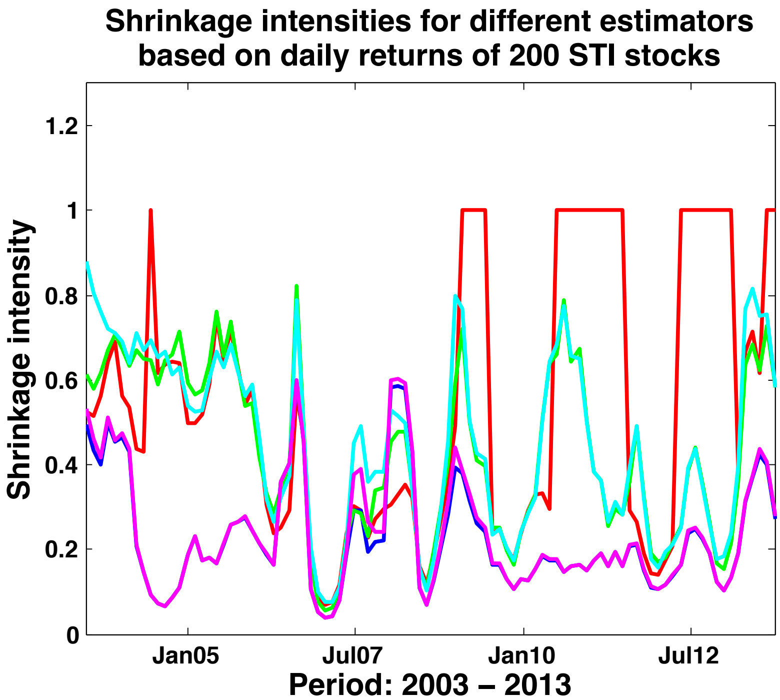
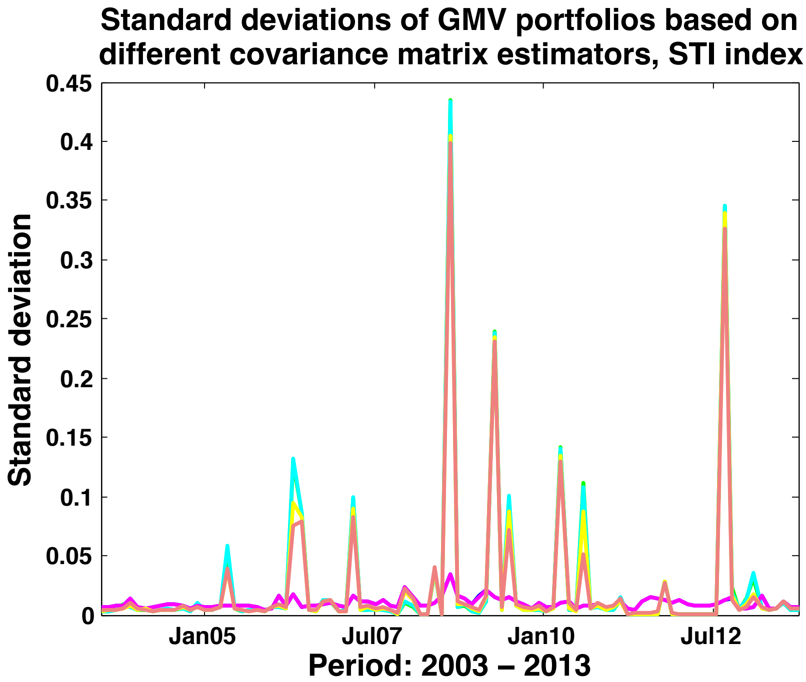


Figure 13: Shrinkage intensities for different covariance estimators (400 stocks); red: shrinkage to constant correlation, blue: shrinkage to identity, green: shrinkage to diagonal m., cyan: shrinkage to market, magenta: shrinkage to two parameteres.



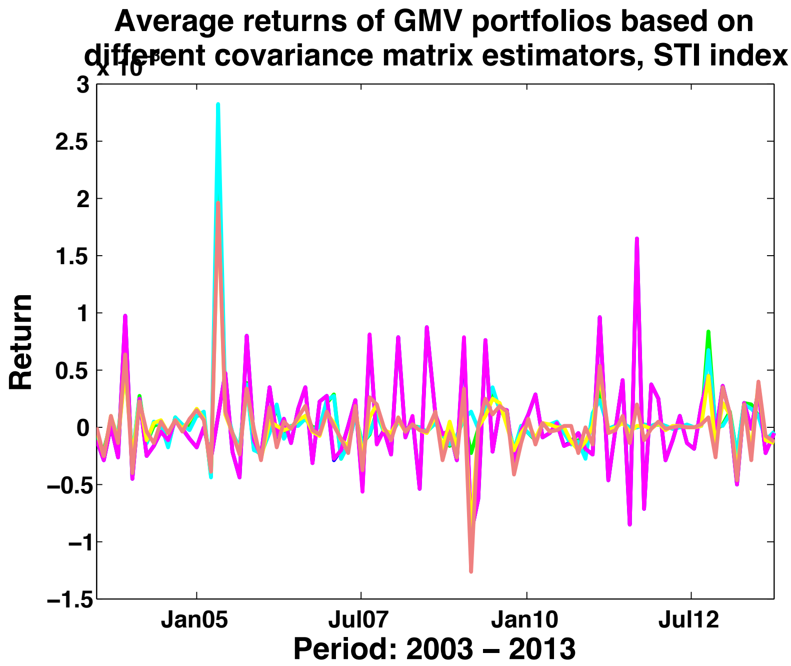


Figure 14: Standard deviations and average returns of 200 STI-stocks GMV portfolio with different covariance estimators; blue: shrinkage to identity, green: shrinkage to diagonal m., cyan: shrinkage to market, magenta: shrinkage to two parameteres, yellow: FFL, coral: SIM.

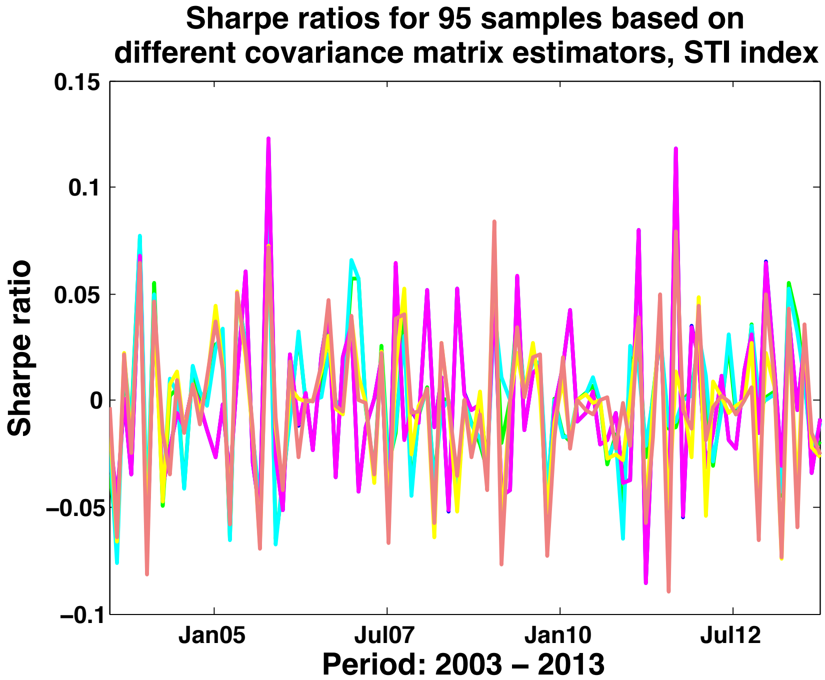
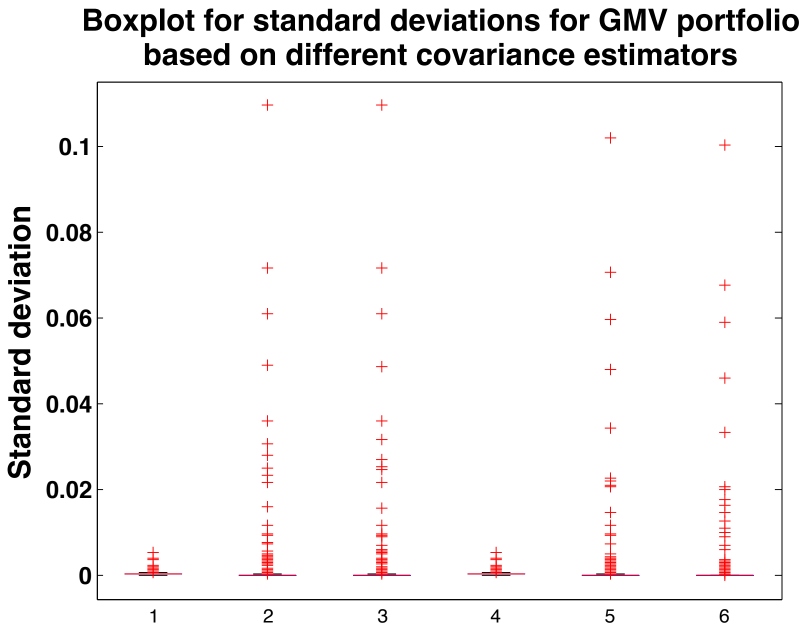


Figure 15: Sharpe ratios for 200 STI-stocks GMV portfolio with different covariance estimators; blue: shrinkage to identity, green: shrinkage to diagonal m., cyan: shrinkage to market, magenta: shrinkage to two parameteres, yellow: FFL, coral: SIM.



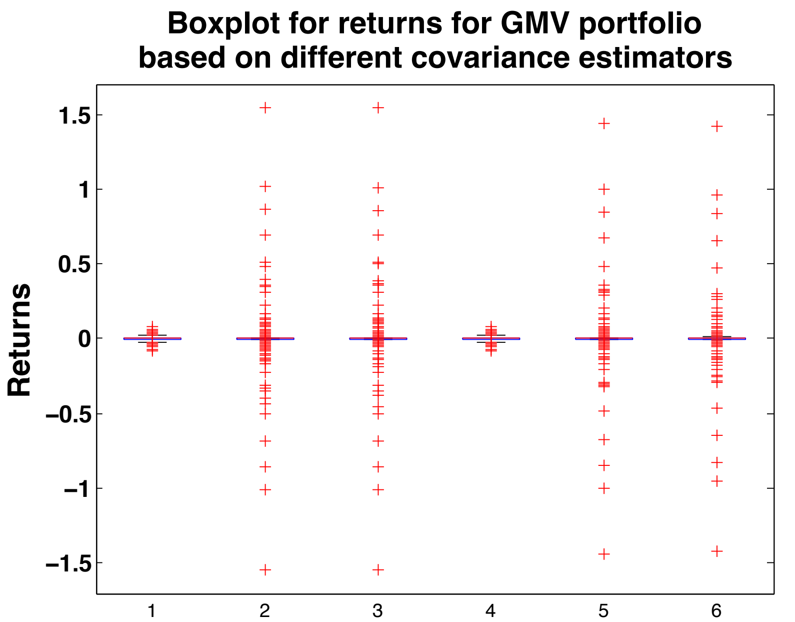


Figure 16: Boxplot for monthly standard deviations and returns of a 200 STI-stocks GMV portfolio with different covariance estimators, 1: shrinkage to identity, 2: shrinkage to diagonal m., 3: shrinkage to market, 4: shrinkage to two parameteres, 5: FFL, 6: SIM

Declaration of Authorship

I hereby certify that this master thesis has been composed by me and is based on my own work, unless stated otherwise. No other person's work has been used without due acknowledgement in this master thesis. All references and verbatim extracts have been quoted, and all sources of information, including graphs and data sets, have been specifically acknowledged.

Natalia Sirotko-Sibirskaya

Berlin, November 19, 2013 Signature: